



## **2.3 Theory of the Firm (Production and Cost)**

### **2.3.1.1 Introduction**

### **2.3.1.2 Organisation of Business Firms**

### **2.3.1.3 Production Function**

- Short Run & Long Run changes in Production

### **2.3.1.4 Production Law & Production Costs in the Short Run**

- Production in the Short Run
  - Law of Diminishing Returns (LDR)
  - Assumptions, Illustration, Observation, Reasons & Relationship between Curves
  - Some Important Points Regarding LDR
- Costs in the Short Run
  - Explicit and Implicit Cost
  - Accounting and Economic Profits
  - Normal Profits
  - Fixed Costs, Variable Costs & Total Costs
  - Average and Marginal Cost Concepts
  - Relationship between MPP and MC & APP and AVC

### **2.3.1.5 Production & Costs in the Long Run**

- Return to Scale
- Economies of scale
- Internal and External Economies and Diseconomies of Scale

### **2.3.1.6 Optimum Combination of Factors**

### **2.3.1.7 Size of Firms**

- Growth of Firms
- Minimum Efficient Scale

### 2.3.1.1 INTRODUCTION

- The current topic tries to examine the **supply side** by studying how the "rational" producer will behave. By "rational", the traditional theory assumes that firms aim to maximize profit.
- Profit is made by firms earning more from the sale of its goods than they cost to produce. In order then to discover how a firm can maximize its profit, we must first consider what determines costs and revenue. In this chapter, we seek to study how production (output) depends on the inputs used, and how costs depend on the amount of output produced.

### 2.3.1.2 ORGANISATION OF BUSINESS FIRMS

- **Production refers to any paid activity, which helps to satisfy wants. It is the process of transferring inputs from human and physical resources into outputs wanted by consumers.** Production ends when consumption begins. Hence, distribution (wholesale and retail trade, transport, etc.) is also considered to be part of the production process.
- The business firm is a production unit, designed to organise raw materials, labour and machines with the goal of producing goods and services.
- Business firms may be organised as follows:

1	<b>Sole proprietorship</b>	Single owner who makes all decisions and bears full responsibility for everything the firm does.
2	<b>Partnership</b>	2 to 20 individuals who own and operate the business jointly.
3	<b>Corporation</b>	Has a legal identity separate from its owners. The corporation enters into contracts that are legal obligations of the corporation but not its owners. The owners are said to have limited liability, for eg. the corporation can be sued but not its owners/shareholders.

- Modern production generally involves large-scale production. There is a need to distinguish between production units of plants, firms and industries:

1	<b>Plants</b>	Production units devoted to the creation of a particular utility or group of utilities. Such a unit will have a distinct geographical site and distinct output or range of output in the form of goods and services.
2	<b>Firms</b>	Business units exerting ownership over one or more plants. The production units they own may be in one particular industry or they may have diversified so that they are engaging in a number of different industries.
3	<b>Industry</b>	A group of firms producing similar or related goods or services for a particular market. Where the goods produced by each firm are substitutes; the firms will be in competition with one another but occasionally are found to be cooperating.

### 2.3.1.3 PRODUCTION FUNCTION

- The production function expresses the technical relationship between the amount of factor inputs, and the output generated per period of time. The inputs used to produce the output may be categorized into land, labour, capital and entrepreneurship.
- The factor inputs can be combined in many ways to produce the same amount of output:  $TPP = f(\text{Land, Labour})$ , which merely states that the total output, or the total physical product (TPP) depends on the amount of land and labour used.

#### 2.3.3.1 Short Run & Long Run Changes in Production

- To change the output produced, a firm changes the amount of resources used. To increase the output, a firm would increase the amount of resources used.
- However, the important role played by time in the production process should be considered here. When a firm needs to increase production in a hurry, it will only be able to increase the quantity of certain factors.
- The distinction we are making here is between fixed and variable factors. **A fixed factor is an input that cannot be increased within a given time period** (eg. buildings). **A variable factor is one that can be increased within the time period considered** (eg. labour, raw materials).
- The distinction between fixed and variable factors allows us to distinguish between the **short run** and the **long run**.

- The short run is the time period so short that the firm is unable to vary all of its factors of production. It is the time period during which at least one factor of production is fixed. In the short run, output can only be increased by using more variable factors: Thus, a manufacturer may increase output in the short run by using more raw materials, more fuel, more tools and more labour, and it will have to make do with its existing buildings and most of its machinery.
- On the other hand, the **long run** is the time period long enough to allow the firm to vary all its factor inputs. Hence, it will be able to vary its plant size, install new machinery as well as all other factors of production in the long run. In addition, in the long run, new firms may enter the industry while others may leave the industry.

#### **Note!**

- ⊕ The short run does not correspond to any particular time period, for example, like a specific number of months or years.
- ⊕ The actual length varies from industry to industry and from firm to firm. It might take an electricity firm two years to build a new power station, but only a few months to open a new food court if an existing building can be bought, converted and decorated.

### **2.3.1.4 PRODUCTION LAW AND PRODUCTION COSTS IN THE SHORT RUN**

#### *2.3.1.4.1 Production in the Short Run*

- To increase output in the short run, the firm can only increase the quantity of variable factors. As the firm keeps increasing the quantity of variable factors while the quantity of fixed factors remains constant, it is combining fixed and variable factors in changing proportions. The behaviour of the inputs in relation to output is expressed in the Law of Variable Proportions or the **Law of Diminishing Marginal Returns**.

##### *2.3.1.4.1.1 Law of Diminishing Returns (LDMR)*

- The **Law of Diminishing Marginal Returns** states that as more and more units of the variable input(s) are added to a given amount of a fixed input, there will come a point where each additional unit of the variable input will add less to total output than previous unit.
- Illustration: Take the case of a wheat farm. Assume the fixed factor is land and the variable factor is labour. In addition, the following conditions must be held:
  - The state of technology is assumed to be constant. This means that any increase in output is due to the change in the employment of variable factors.
  - Labour is homogenous and all workers are equally skilled. This ensures when the additional output falls upon the employment of an additional worker, it is not because he is inefficient.
- In order to increase its production of wheat in the short run, the farmer adds more and more workers to his fixed plot of land.

- Table 1 shows the hypothetical production function in a table form. The first two columns show how total wheat output varies as extra workers are employed on a fixed amount of land.

**Table 1: Wheat Production in the Short Run**

Quantity of Labour ( $Q_L$ )	Total Physical Product (TPP)	Average Physical Product (APP)	Marginal Physical Product (MPP)
0	0	-	3
1	3	3	7
2	10	5	14
3	24	8	12
4	36	9	4
5	40	8	2
6	42	7	0
7	42	6	-2
8	40	5	

*Definition of concepts:*

- **Total product** (or total physical product, TPP) is the total amount of output produced during a given period of time where all the factors of production are employed.
- **Marginal product** (or marginal physical product, MPP) of the variable factor refers to the additional output gained by the employment of one more unit of the variable factor. Mathematically,

$$\text{MPP} = \Delta \text{TPP} / \Delta Q_v,$$

where  $Q_v$  is the quantity of variable factor.

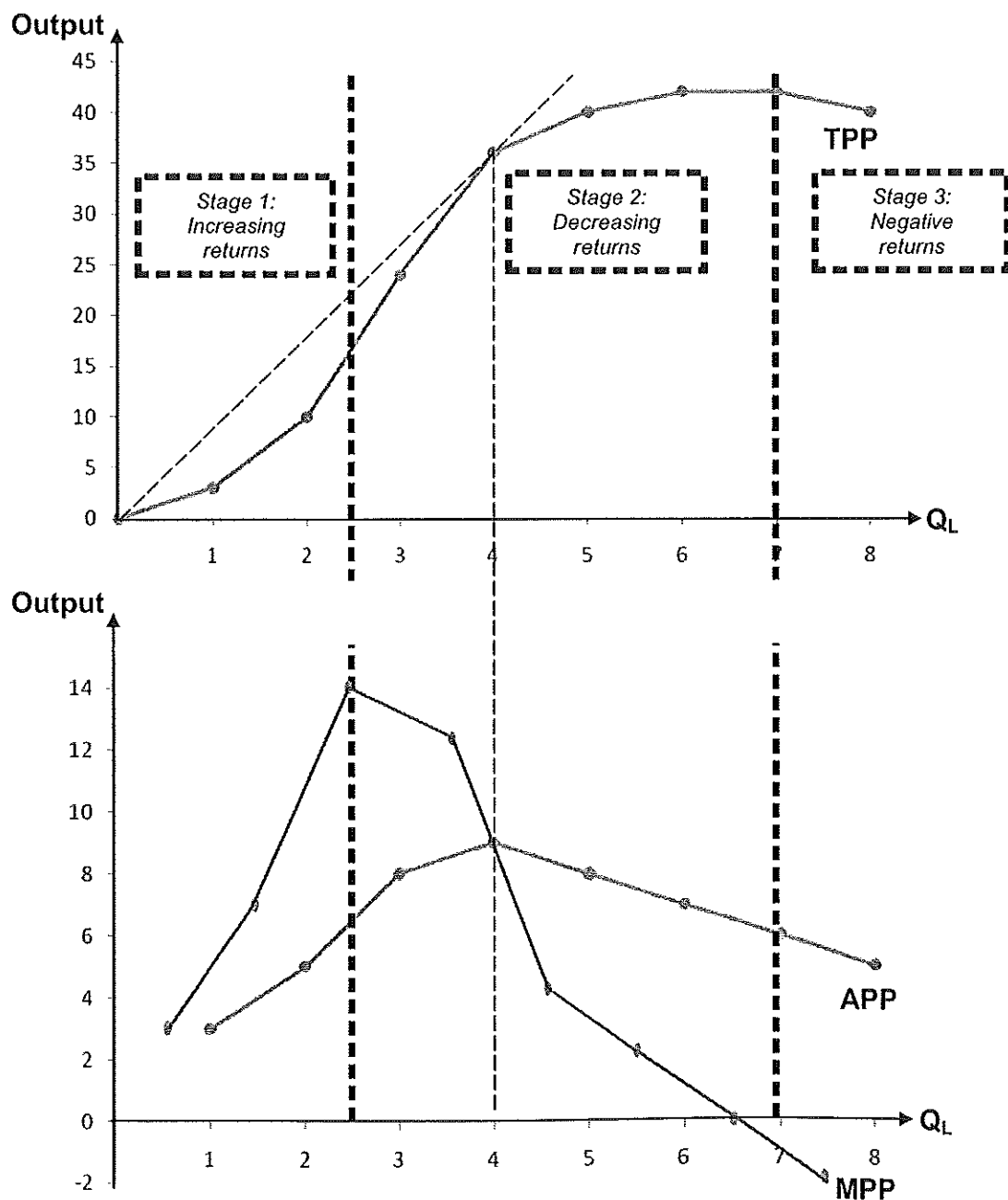
- **Average product** (or average physical product, APP) is the total output (TPP) per unit of the variable factor in question. Mathematically,

$$\text{APP} = \text{TPP} / Q_v.$$

- From Table 1, the following observations can be made:
  - With nobody working on the land, output will be zero.
  - As the first three farm workers are taken on, the total wheat output initially rises more and more rapidly (Column 1). In other words, the addition to total output with the addition of the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> worker (MPP) increases from 3 to 14.

- Upon the employment of the 4<sup>th</sup> worker, diminishing returns set in. Notice that MPP drops to 12.
- This fall in MPP continues when more workers are added. Eventually when the 7<sup>th</sup> worker adds nothing to TPP, the land has yielded as much as it can. Any more workers employed after that are likely to get in each other's way, resulting in the eight workers producing less than the seven (negative MPP).
- Diagram 1 is the graphical expressions of the above production function:

**Diagram 1: TPP, APP and MPP (based on Table 1)**



- Explanation of the Diagram 1:
  - **Stage 1: Stage of increasing MPP**
    - Initially, land is relatively plentiful and is under-utilised; with only one or two workers, efficiency (productivity) is low, as the workers are spread too thinly.
    - With more workers employed, specialisation and division of labour becomes possible. Land is more intensively utilised by labour. The factor combination becomes more efficient and productivity increases. Notice that slope of the TPP curve increases which is consistent with a rising MPP.
  - **Stage 2: Stage of diminishing but positive MPP**
    - In stage 2 the slope of TPP declines, suggesting a declining MPP
    - Land, the fixed factor, becomes relatively scarce and is over-utilised. There is the problem of over-crowding as labour is now under-utilised, with some labour idle part of the time. The factor combination becomes less efficient, with the productivity of labour falling.
  - **Stage 3: Stage of negative MPP**
    - The factor combinations moves further away from the optimum as more workers are added, leading to a stage when the employment of additional units of labour would result in a fall in total output (the MPP of the 8<sup>th</sup> worker is negative).

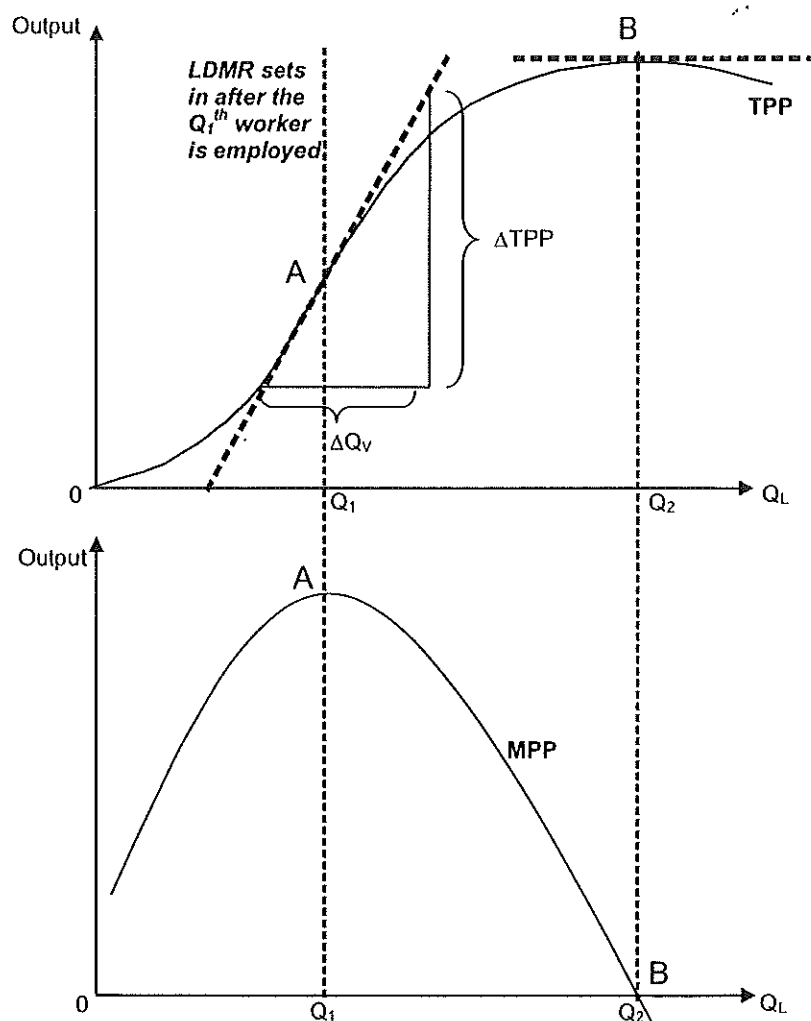
#### 2.3.1.4.1.2 Some Important Points Regarding LDMR

- The law applies only if at least one factor is held fixed. If both or all the factors of production can be varied it is no longer the short run but is the long run.
- The law is merely technical, stating physical relationships for it only looks at the physical input vs. physical output, ignoring the element of cost and revenue. So when considering how many workers to employ, LDMR does not determine the decision.
- The technique of production must be held constant for if technology changes, then the change in output is due to technological changes and diminishing returns may not set in.

2.3.1.4.2 Relationships between TPP, APP and MPP  
(For simplicity, the variable factor assumed here is labour)

2.3.1.4.2.1 Relationship between TPP and MPP

**Diagram 2: TPP and MPP**



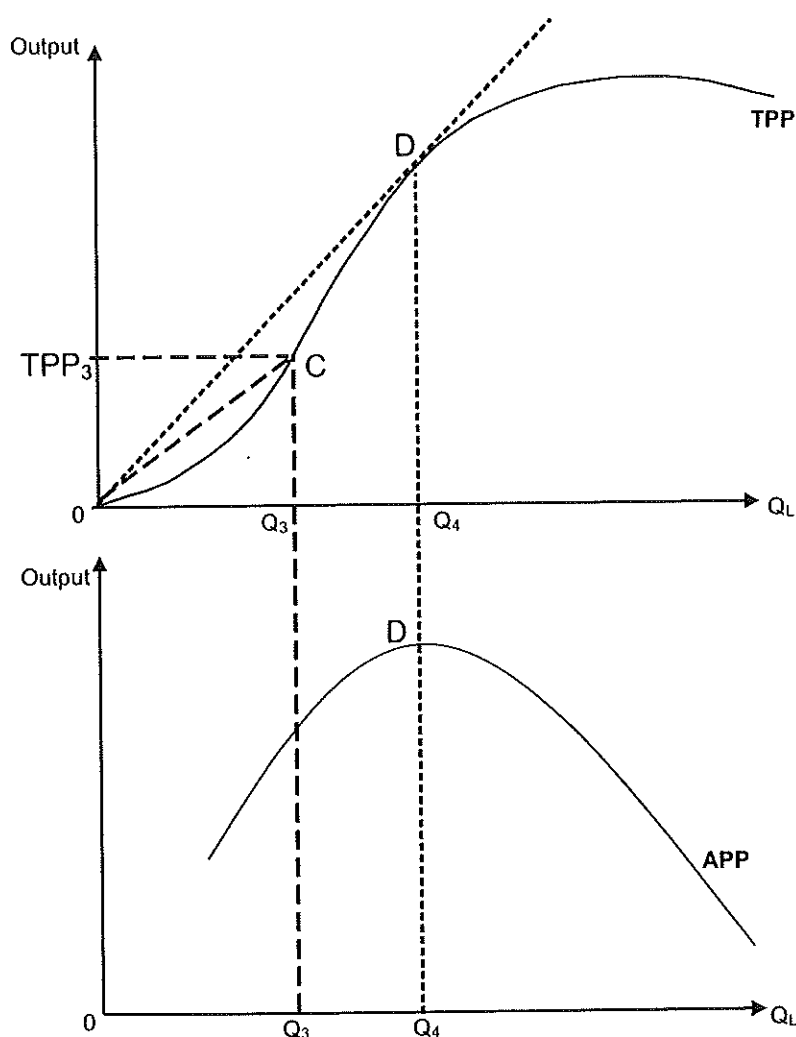
- Graphically, MPP is derived from the slope (gradient) of the TPP curve between two points.
- With reference to Diagram 2, between 0 and  $Q_1$  number of workers employed, MPP rises at first and correspondingly, the slope of the TPP curve gets steeper. MPP reaches a maximum at point A. At this point the slope of the TPP is the steepest.
- Diminishing marginal returns set in after point A when MPP starts to fall, but it is still positive (TPP becomes less steep).



- As long as MPP is greater than zero (ie. employment of new workers add to total output), TPP will go on rising. At point B (when the  $Q_2^{\text{th}}$  worker is employed, TPP is at a maximum, and MPP is zero (slope of TPP is zero), implying that the addition of that worker adds nothing to output. Beyond point B, TPP falls as MPP is negative.

#### 2.3.1.4.2.2 Relationship between TPP and APP

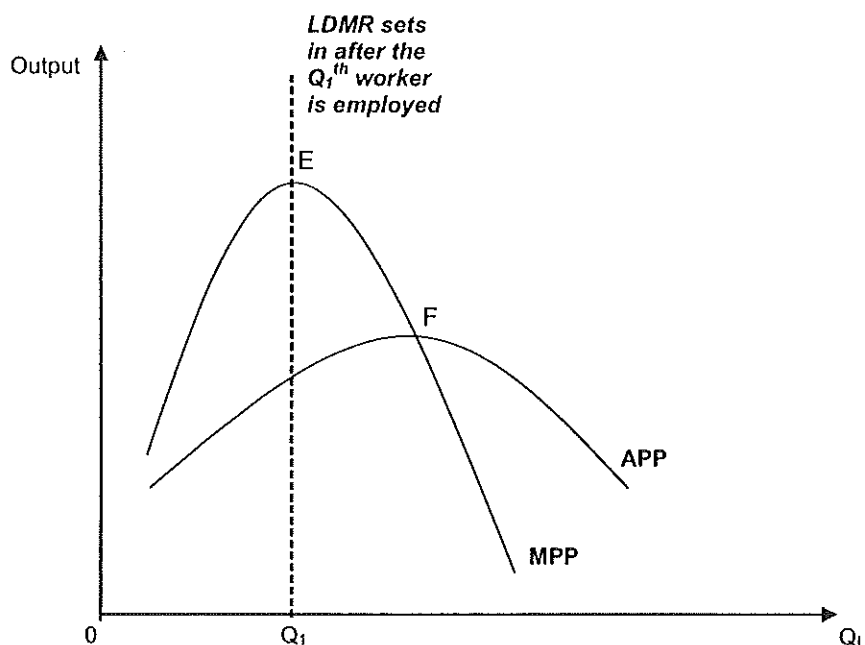
**Diagram 3: TPP and APP**



- Since  $APP = TPP / Q_L$ , it follows that APP is given by the slope of the ray from the origin to the relevant points on the TPP curve. Referring to Diagram 3, when  $0Q_3$  workers are used, the APP is equal to the slope of the ray  $OC$ , ie.,  $CQ_3/0Q_3$ .
- It is clear from Diagram 3 that APP reaches its maximum where the ray from the origin is tangent to the TPP curve, ie., at the point D where  $0Q_4$  workers are employed.

#### 2.3.1.4.2.3 Relationship between APP and MPP

**Diagram 4: APP and MPP**



Referring to Diagram 4,

- APP rises at first. APP continues to rise as long as the MPP is greater than the APP; ie., the addition to output from the last worker is greater than the average output of the previous workers. Hence, the MPP pulls the APP up. This continues beyond point E and until point F. Hence, even when MPP is falling, APP keeps rising as long as MPP is still above APP.
- Beyond point F, MPP is below APP. This means that new worker adds less to total output than the average output of the previous workers. This pulls the average down and APP falls.

### Learning Tip!

To help you understand the relationship between average and marginal better, consider the case of a class with 10 people in it and their average age is 20. Now if a 20-year-old enters the class (the marginal age), the average age remains the same. If a 50-year-old enters the class the average age rises. When a child of 10 is to enter the class, this will pull down the average age of the people in the class. Hence,

- If the marginal equals the average, the average will not change.
- If the marginal is above the average, the average will rise.
- If the marginal is below the average, the average will fall.

#### 2.3.1.4.3 Costs in the Short Run

- We have seen how output varies with inputs in the short run. We now use this information to show how costs vary with the amount a firm produces. A firm needs to know what the costs are at each level of output to know how much to produce. Costs depend on the quantities of factor inputs used and the prices of these inputs.
- The term 'costs' is used differently by economists and accountants.

##### 2.3.4.3.1 Explicit and implicit costs

- Economics use the concept of **opportunity cost** when measuring costs.
- Consider the example of a chicken rice seller called Mr Lee. The costs he incurs in a year in terms of the price he has to pay to the land, labour and capital resources he employs is shown below:

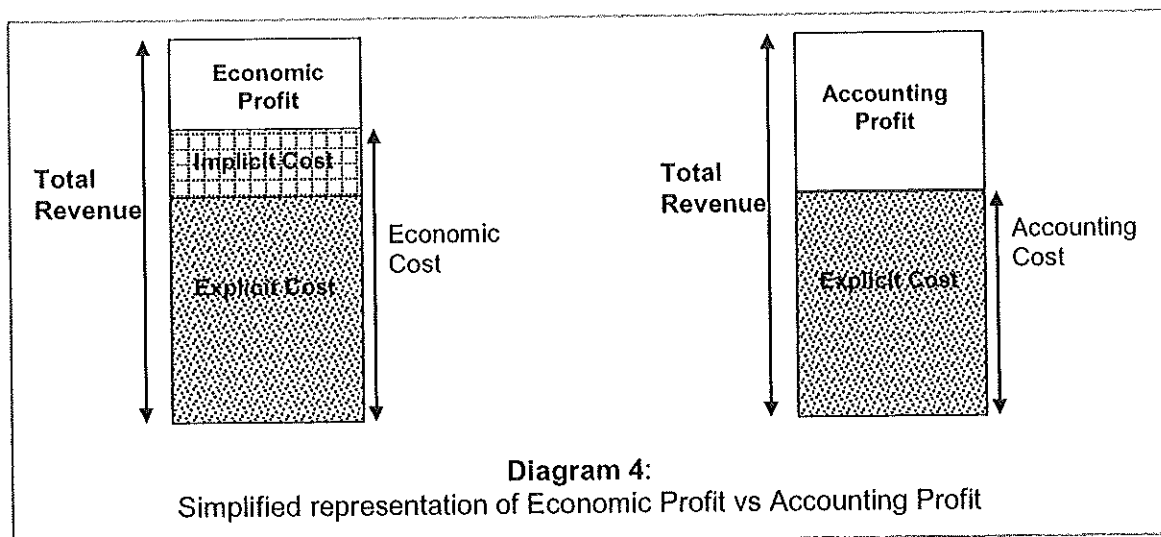
Hired labour	\$10,000
Utilities	\$ 2,000
Supplies	\$ 5,000
Depreciation of equipment	\$ 2,000
Taxes	\$ 1,000
<b>Total</b>	<b>\$20,000</b>

- The total cost is \$20,000: But this is only so for the accountant.
- For an economist, the above merely represents **explicit costs**. **Explicit costs refer to production costs which are explicitly incurred. They are what the firm pays out to purchase or hire factor inputs.** These costs are called explicit costs as they involve direct money payment by the firm.
- However, a firm does not need to pay to use the factors which are owned by the firm itself. However, opportunity costs are still incurred as these factors could earn income for the firm from some alternative use, either within the firm or hired out to some other firm. These opportunity costs are known as **implicit costs**. **Implicit costs are costs that do not involve any direct money payment to a third party, but which nevertheless involve a sacrifice of some alternative.** Examples include:
  - The salary of \$12,000 Mr Lee could have earned as a clerk in an office instead of having the food stall.
  - If he invests \$10,000 from his savings in his business, the opportunity cost is the interest foregone. Assuming that the bank pays an interest of 10%, the foregone interest is \$1,000.
  - Imputed rent of \$10,000 on a shophouse owned by Mr Lee which he uses for his stall.

Total implicit costs = \$23,000

#### 2.3.4.3.2 Economic Profit vs Accounting Profit

- Total Profit = Total Revenue - Total Cost
- Accounting Profit = Total Revenue - Explicit Cost
- Economic Profit = Total Revenue - (Explicit & Implicit Costs)



- Assuming that the total revenue earned by Mr Lee from the sale of chicken rice in a given year = \$80,000.
- Accounting Profit = \$80,000 - \$20,000 = \$60,000.
- But to the economist, it is less since the total cost of production is the sum of the implicit and the explicit costs incurred from the employment of all the factors of production → Economic Profit = \$80,000 - \$20,000 - \$23,000 = \$37,000.

#### 2.3.1.4.3.2.1 Normal Profit

- One element of cost is the opportunity cost to the owners of the firm of being in business. This is the minimum return the owners must make on their capital in order to prevent them from eventually deciding to close down and perhaps to move into some alternative business. It is a cost because, just as with wages, rent etc., it has to be covered if the firm is to continue producing. This opportunity cost to the owners is sometimes known as **normal profit**.
- Normal profit is considered an implicit cost because it is an opportunity cost of entrepreneurship to the firm. It is the payment to Mr Lee to compensate him with a minimum amount of profit for bearing risks. If not, he would rather work for somebody and not be the owner or switch to other businesses.

<b>Table 2: The Opportunity Cost of each Factor of Production</b>	
<b>Factor of Production</b>	<b>Opportunity Cost</b>
Land	Rent
Labour	Wages
Capital	Interest
Entrepreneurship	Normal Profit

#### 2.3.1.4.3.3 Short Run Costs

- The costs of production of a firm depend on the factors of production it uses. The more factors it uses, the greater will its costs be. More precisely, this relationship depends on two elements:
  - The productivity of the factors. The higher the productivity, the fewer the amount of factors that will be needed to produce a given level of output and the lower will be the costs of producing that output.
  - The price of the factors. The higher their price, the higher will be the costs of production.
- Since in the short run, some factors are fixed while others are variable, a firm's total cost of production may be divided into **fixed and variable costs**.

#### 2.3.1.4.3.3.1 Fixed costs

- Since fixed factors are fixed in supply in the SR, the total costs of such factors do not vary with output, as shown in Diagram 5 by the curve TFC (total fixed cost). These will be present whether the firm produces 20 units of the output, 100 units or 0 unit.
- For example, the firm must still make rental payments on the lease of property and interest payments on any loans taken out. Other examples include depreciation on plant and equipment, property taxes and insurance.

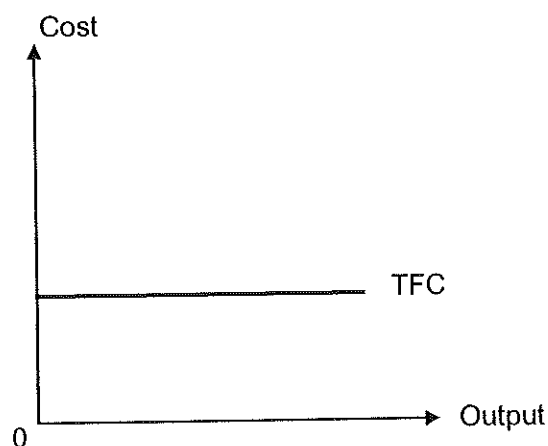
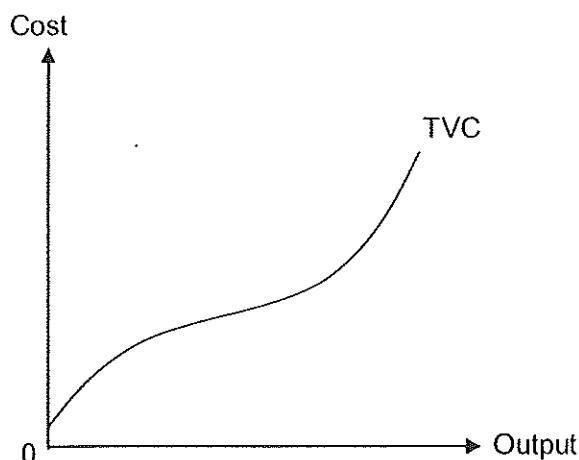


Diagram 5: Total Fixed Cost Curve

#### 2.3.1.4.3.3.1 Variable costs

- Variable costs are those that vary with output. Total variable costs increase as more output is produced and decrease when less is produced, as shown in Diagram 6. When output is zero, variable costs are zero.
- Examples include payments for wages, raw materials, fuel and power, and transport costs.

Diagram 6: Total Variable Cost Curve



#### 2.3.1.4.3.4 Total Costs

- In the short run, the total cost of production is the sum of the total fixed costs and the total variable costs.

$$TC = TVC + TFC$$

(Note that costs here includes both explicit and implicit costs.)

Table 3: Cost schedules for a firm producing at different output levels

Quantity	TFC (\$)	TVC (\$)	TC (\$)	MC (\$)	AC (\$)	AFC (\$)	AVC (\$)
0	55	0					
1	55	30					
2	55	55					
3	55	75					
4	55	105					
5	55	155					
6	55	225					
7	55	315					
8	55	425					

Diagram 7 shows the graphical illustrations of the TFC, TVC and TC schedules.

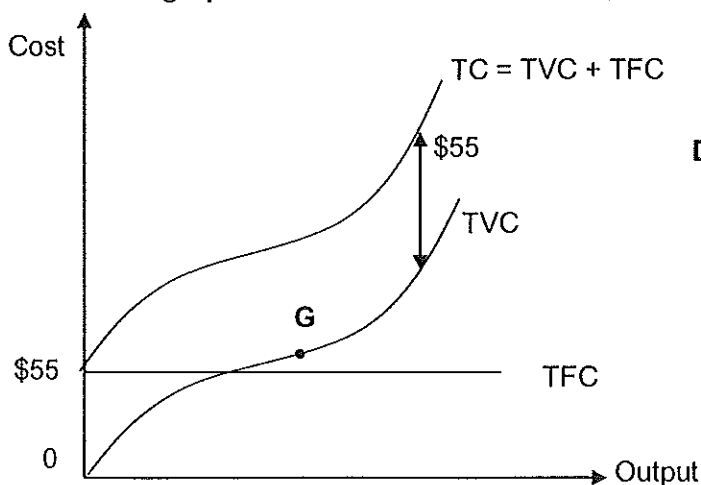


Diagram 7: TFC, TVC & TC

- Since TFC is fixed at \$55 and does not vary with output, it is shown by a horizontal straight line in Diagram 7.
- When output is zero, no variable factor is used and the TVC equals zero. Thus the TVC curve starts at the origin. The shape of the TVC curve may be explained by the *Law of Diminishing Marginal Returns*.
  - Initially, during the stage of increasing returns, TVC rises less and less rapidly as each additional variable factor is used. This corresponds to the portion of the TPP curve that rises more and more rapidly (refer to Stage 1 in Diagram 1). This refers to the case of a firm with a certain amount of fixed factors, and when more and more workers (variable factors) are employed initially, efficiency increases and output increases more and more rapidly as the workers are able to specialise and the fixed factor is used more fully.
  - However, when output is increased beyond **point G** in Diagram 7, diminishing returns set in. Additional workers produce less and less extra output. As such, the extra units produced by these workers will be costing more and more in terms of wage costs. Thus, the TVC curve gets steeper and steeper as TVC increases more and more rapidly. Hence, this corresponds to the TPP curve that rises less and less rapidly (refer to Stages 2 & 3 in Diagram 1).
- Since TC is the addition of TFC and TVC, the TC curve is simply the TVC curve shifted vertically upwards by the amount of TFC (in this case, \$55).

#### 2.3.1.4.3.4.1 Average and Marginal Costs Concept

- **Average Cost (AC)** is cost per unit of output produced. It is calculated by dividing total cost by the output (Q). That is,

$$AC = TC/Q$$

- Like total cost, AC can be divided into two components – average fixed cost and average variable cost. That is,

$$AC = AFC + AVC$$

**Average Fixed Cost (AFC)** is the total fixed cost per unit of output. That is,

$$AFC = TFC/Q$$

**Average Variable Cost (AVC)** is the total variable cost per unit of output. That is,

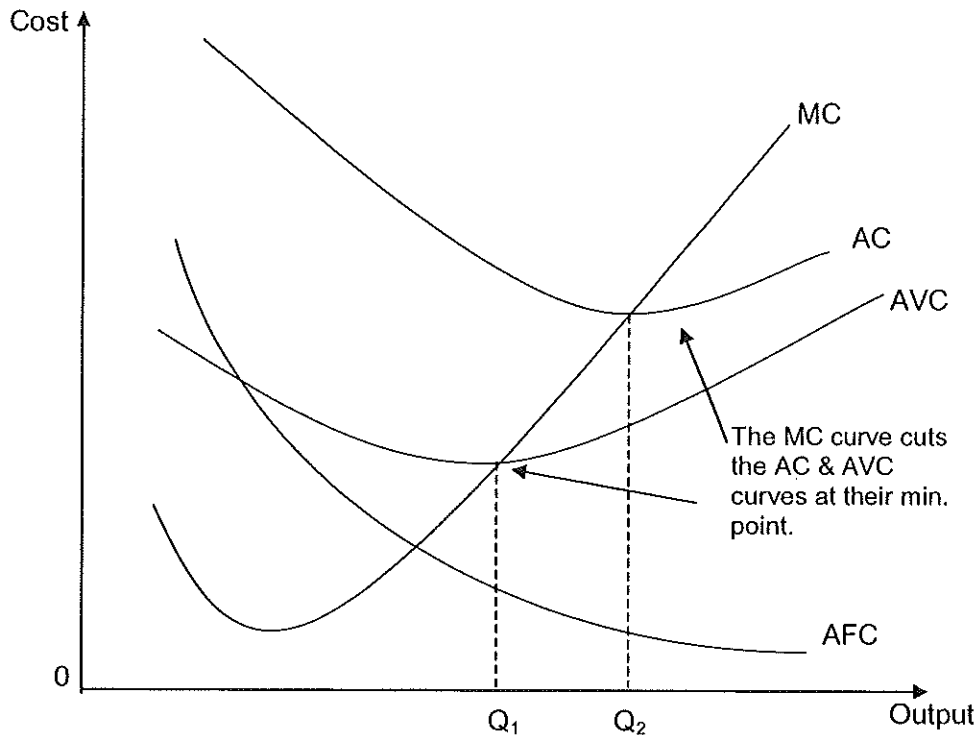
$$AVC = TVC/Q$$

- **Marginal Cost (MC)** refers to the addition to total cost as a result of producing an additional unit of output. That is,

$$MC = \Delta TC/\Delta Q \text{ or } \Delta TVC/\Delta Q$$

- The slope of the TC curve measures MC. Since TFC is constant regardless of the level of output produced, the change in TC is exactly equal to the change in TVC. Hence, MC will also be the slope of the TVC as the slope of both the TC and TVC is equal.
- The MC, AFC, AVC and AC curves are shown in Diagram 8.

Diagram 8: AFC, AVC, AC & MC



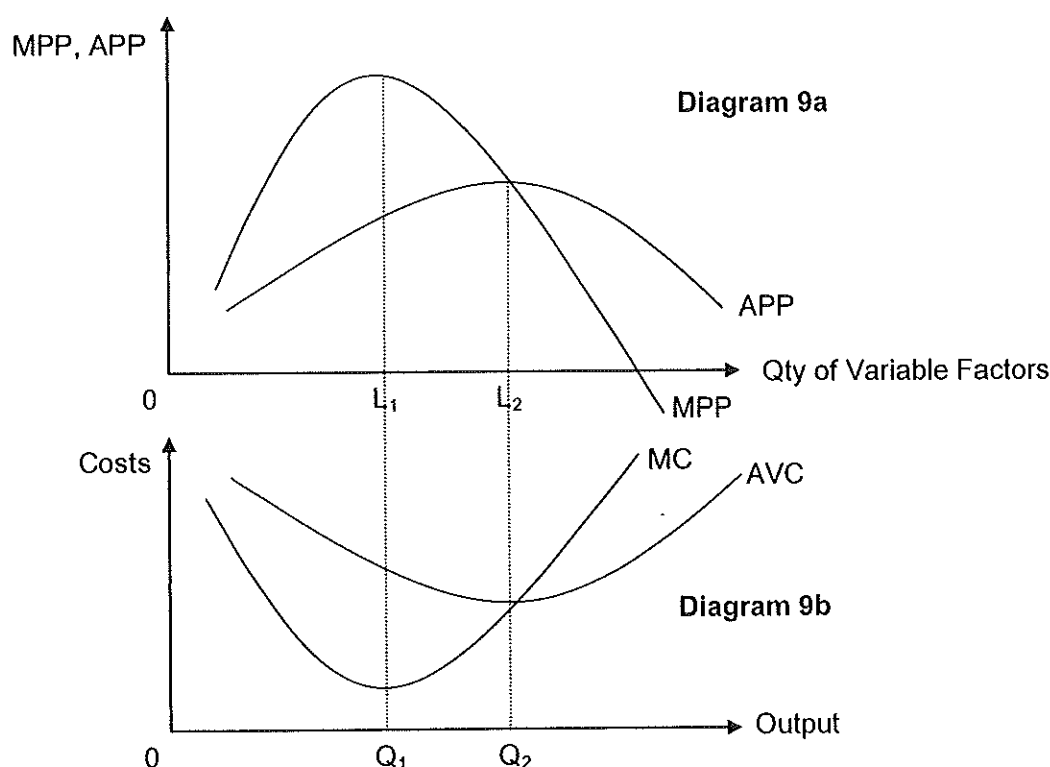
- **Average Fixed Cost (AFC)** falls continuously as output increases, since TFC does not vary with output. As more output is produced, total fixed costs are being spread over a greater and greater output.
- **Average Variable Cost (AVC)** curve is also affected by the *Law of Diminishing Marginal Returns* and its shape depends on the shape of the APP curve. As the APP of workers rises, the average labour cost per unit of output (AVC) falls. Thereafter, as APP falls, AVC must rise. Therefore AVC curve takes on a U-shape, falling initially, reaches a minimum at output level  $Q_1$ , then rises continuously.
- **Average Cost (AC)** curve is merely the vertical sum of the AVC and AFC curves [ $AC = AFC + AVC$ ]. AC is also U-shaped. That is, AC decreases as output is increased until it reaches a minimum point at output level  $Q_2$ , then as output is further increased, AC rises.



#### Relationship between AVC, AFC and AC when output rises:

- Over the range of output  $0Q_1$ , both AFC and AVC are falling, which also indicates that AC must be falling.
- Over the range of output  $Q_1Q_2$ , even though AVC is rising, AC will not rise because the decrease in AFC exceeds the increase in AVC. Thus AC will continue to fall.
- When the increase in AVC just offsets the decrease in AFC, AC reaches its minimum at output level  $Q_2$ .
- As soon as the increase in AVC is greater than the fall in AFC, AC curve rises.

#### Relationship between MPP & MC and APP & AVC curves



- MC falls at first as output rises, reaches a minimum and then begins to rise when diminishing returns set in. Hence, the MC curve is U-shaped. This is illustrated in Diagram 9b. Initially as more of the variable factor is used, each extra unit of output cost less than the previous units, MC falls.
- This corresponds to the rising portion of the MPP curve as shown in Diagram 9a. This occurs because each additional unit of variable factor employed adds more to output than the previous unit but the additional cost incurred in employing them remains unchanged.
- Beyond a certain level of output ( $Q_1$ ), diminishing returns set in. From this point on, MC rises as MPP falls as additional units of output cost more and more to produce as each additional unit of variable factor employed adds less and less to output than the previous unit but the additional cost incurred in employing them remain unchanged.

- Similarly, initially, as more and more units of variable factors are employed, average productivity is increasing and, so, the APP rises. This corresponds to the downward sloping portion of the AVC curve as shown in Diagram 9b. Beyond a certain level of employment ( $L_2$ ), average productivity begins to fall and this results in the APP sloping downwards. This corresponds to the upward sloping portion of the AVC curve as shown in Diagram 9b.

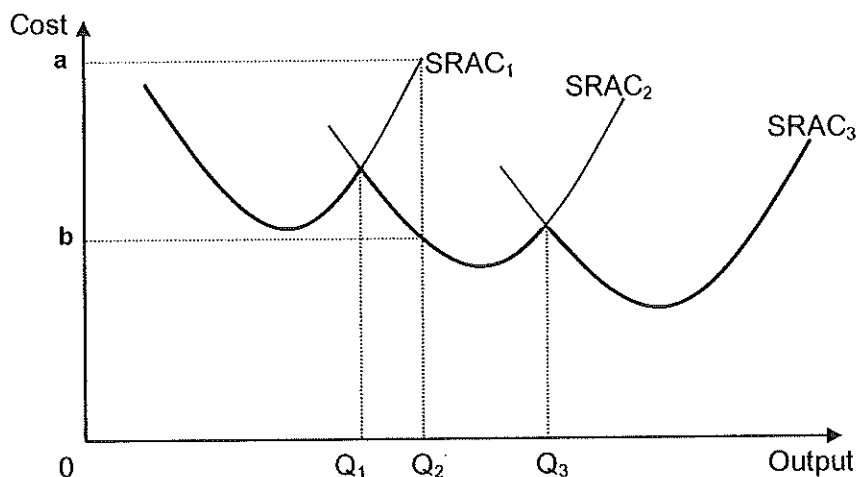
#### Relationship between AC & MC

- The relationship between AC and MC is the same with the general rule that defines the relation between all average and marginal variables.
  - As long as MC is less than AC, AC must be falling since when the addition to TC (MC) is less than the average cost (AC) of the previous units, the production of the new units must pull the average cost down.
  - Similarly, when additions to TC are more than the average cost, the average cost is driven up. Thus if MC is greater than AC (that is, MC lying above AC), AC must be rising. Hence, the MC cuts the AC at its minimum point from below.
  - AVC and MC are similarly related, that is,
    - When MC is less than AVC, AVC will decrease.
    - When MC is more than AVC, AVC will increase.
- Hence, MC also cuts AVC at its minimum point from below.

### 2.3.1.5 PRODUCTION AND COSTS IN THE LONG RUN (LR)

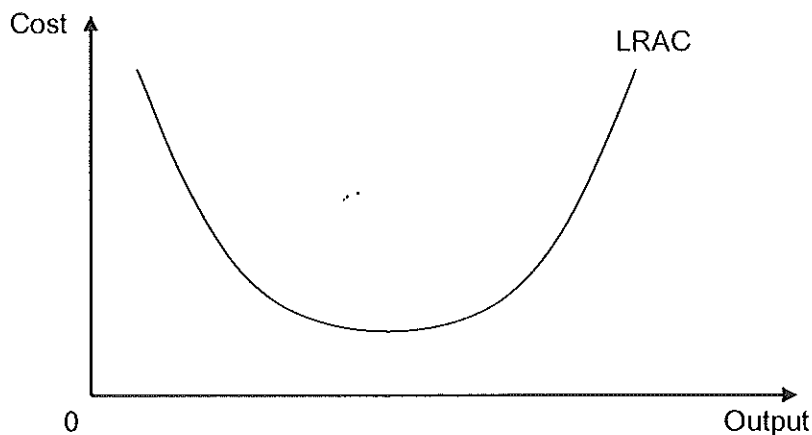
- All factors of production are variable in the LR. A firm may expand its present factory, build new factories in different parts of the country, install new machines and even change its techniques of production in the LR. In the LR, it may make decisions regarding its scale of production, the location of the factories and the methods of production, each of which affect its costs of production.

Diagram 10: Derivation of LR Average Cost from SR Average Cost Curves



- In the long run, the firm can change its plant size to reduce its cost of production. Assume that there are three plant sizes only. Diagram 10 shows 3 SR average cost curves for three plant sizes that can produce successively higher levels of output.
  - Assuming the firm is currently operating *Plant Size 1* (SRAC<sub>1</sub>). For any output below OQ<sub>1</sub>, the firm will use *Plant Size 1*. If output expands to OQ<sub>2</sub>, the firm will continue to operate *Plant Size 1* in the short run as the firm cannot alter the plant size it uses in the short run. Its average cost of producing OQ<sub>2</sub> in the short run will be 0a.
  - In the long run, the firm will be able to alter its plant size and move to *Plant Size 2*. This will lower the average cost of producing OQ<sub>2</sub> in the long run to 0b. Between output Q<sub>1</sub> and Q<sub>3</sub>, the firm will use *Plant Size 2*.
  - Beyond output OQ<sub>3</sub>, *Plant Size 3* will be used in the long run to lower average cost.

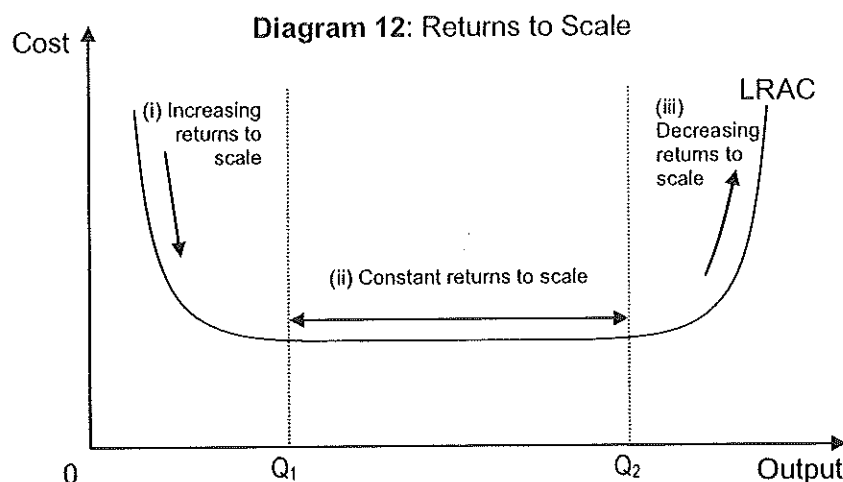
Diagram 11: Long Run Average Cost Curve



- The LRAC is derived from a number of short run average cost curves. The bold curve in Diagram 10 gives the LRAC curve. The LRAC curve represents the lowest cost of producing each level of output when the firm is free to change its plant size. It is also known as the *planning curve*.
- If we assume that there are an infinite number of plant sizes available to the firm in the long run, the LRAC curve will be a smooth one as shown in Diagram 11.

#### 2.3.1.5.1 Returns to Scale

- In the SR, plant size is fixed and so only one scale of production exists. It follows that in the long run, a firm considers all factors of production variable. It can grow or it can shrink in size. In the long run, there are 3 possible effects on output from an increase in the level of use of all factors.
  - i) **Increasing returns to scale** means a proportionate increase in inputs will lead to a more than proportionate increase in output. For example, a 10% increase in all inputs results in a more than 10% increase in output. As a result, average (unit) costs of production decrease. The firm enjoys **internal economies of scale**.
  - ii) **Constant returns to scale** means a proportionate increase in inputs will lead to a proportionate increase in output. For example, an increase in all inputs by 10% will lead to a 10% increase in output. As a result, average (unit) costs of production remain constant.
  - iii) **Decreasing returns to scale** means a proportionate increase in inputs will lead to a less than proportionate increase in output. For example, an increase in all inputs by 10% will lead to a less than 10% increase in output. As a result, average (unit) costs of production rises. The firm suffers **internal diseconomies of scale**.



**Note:** The term 'to scale' is used when all inputs are increased by the same proportion. The long run production theory, **Law of Returns to Scale**, assumes a fixed factor proportion because changes in output will be due only to the scale of production and not because of factor proportions.

#### 2.3.1.5.2 Economies of Scale

- When the long run cost per unit of output decreases as the scale of production is increased, we say that the firm is experiencing economies of scale. Hence, economies of scale are the cost savings enjoyed from the growth in the production scale of the firm or the growth of the whole industry.

#### 2.3.1.5.3 Internal Economies of Scale:

- These are advantages which reduce the long-run average cost as the **firm expands its own scale of production**; that is, its own size.

##### a) Technical Economies

##### i) **Specialization and Division of Labour**

If the scale of production is large, workers can do simple repetitive work as production can be broken down into simpler, more repetitive tasks. With this specialisation and division of labour, less time will be wasted by workers switching from one operation to the next and supervision becomes easier. There is also scope for specialisation of machinery. Thus productivity of the firm may be considerably increased, lowering the unit costs of production.

ii) **Indivisibilities**

Some inputs, especially machinery, tend to be indivisible; that is, they are of a minimum size and cannot be bought in smaller units. Example, a small-scale farmer would not be able to make full use of the combine harvester and it only becomes economical to use when output is large. When output is large, cost of such equipment per unit of output falls considerably. Large firms can invest in capital equipment and operations such as computer system or assembly lines that will not be optimally used by a small firm.

iii) **The 'Container Principle'**

According to this principle, any capital equipment that contains things, e.g., oil tankers or furnaces will cost less per unit of output the larger its size. The cost of the container depends mainly on the surface area while the output depends mainly on the volume. Since large containers have a bigger volume relative to the surface area than do smaller containers, large containers tend to be more cost effective. E.g., an eightfold increase in capacity can be gained by a fourfold increase in the container surface area and hence an approximate fourfold increase in cost.

- The above listed economies of scale arise as a result of the individual workplace, factory/plant or machine being large in size. The following economies of scale that arise as a result of the firm being large (for instance, a firm with many factories.)

b) **Organizational / Managerial Economies**

Cost savings may be made when the employment of the best managers is spread over a large output. In addition, in the case of a large firm, individual plants may specialise in particular functions and thus increase efficiency and decrease cost per unit of output. Centralised administration for instance, after a merger between two firms, also tend to be cost effective.

c) **Spreading Overheads**

Certain expenditures like those on R&D are only cost effective when the firm is large. Moreover, only a large firm is able to afford a research laboratory that undertakes research to increase productivity and initiate cost savings. Hence, the greater the firm's output, the more the overhead costs are spread.

d) **Commercial / Marketing Economies**

i) **Purchase of inputs**

A large firm can usually purchase inputs at a lower price than the small firm as it buys in bulk. Bulk discounts substantially reduce the costs of each unit of material purchased.

ii) **Sale (Marketing) of output**

Distribution on a larger scale (bulk distribution) may enable the firm to fully utilize larger warehouses, delivery vehicles and even set up its own distribution fleet. The large firm is also able to promote sales through bulk advertising. The small firm may not be able to advertise since advertising cost per unit of output will be too high.

**e) Financial Economies**

Larger firms are often able to get bank loans at lower rates of interest than smaller establishments. This is because they are considered to be less risky customers than the small firms. Moreover, large firms like joint-stock companies can increase their issue of shares and debentures to get more capital to expand production.

**f) Risk-bearing Economies**

Risks are spread out by diversifying production. Hence, a firm producing computers may diversify into the textile and food industry. Hence, if a particular good is not selling well, the loss may be offset by the profits from another product.

*(Note: This type of economies of scale helps to minimize losses but does not reduce average costs of production)*

**2.3.1.5.3 Internal Diseconomies of Scale:**

When a firm increases beyond a certain size, costs per unit of output may start to increase, that is, internal diseconomies of scale may set in.

These may arise as a result of:

**a) Problems of a Bureaucratic Structure**

As a result of the large size of the firms, problems of coordination may arise. As large organisations tend to have many specialised departments like sales, production and personnel, the task of coordinating their activities may become more difficult. Also, the larger firms tend to be bogged down by rules and regulations and decision-making is slowed down by the need to go through "proper channels". Work efficiency is reduced by excessive paper work.

**b) Problems of Communication**

As lines of communication get longer, there may be a lack of personal involvement by management. It may be difficult for the management to ensure that everyone gets the necessary information as well as give constructive feedback to the workers to boost their efficiency. Large organisations also do not lend easily to effective control and it would be difficult to ensure that everyone does what they are supposed to do. Insufficient supervision may result in deterioration in the quality of the product.

**c) Absence of Personal Element**

There would be a higher occurrence of feelings of alienation amongst the workers as they will feel they are an insignificant part of a large organisation. Moreover, due to the high degree of specialisation, workers may get bored with their repetitive tasks. Poorly motivated workers may produce shoddy work. As productivity falls, costs per unit of output tend to rise.

**d) Marketing Diseconomies**

Firms which already have a significant market share can only increase market share by spending proportionately more on certain types of marketing expenditure. For example, if a firm has a market share of 75%, an increase in advertising expenditure of 100% is unlikely to increase market share to 100%.

Advertising expenditure per unit (1%) of additional market share rises as market share rises. Thus, beyond a certain level of market share, firms incur marketing diseconomies that will raise their long run AC and MC.

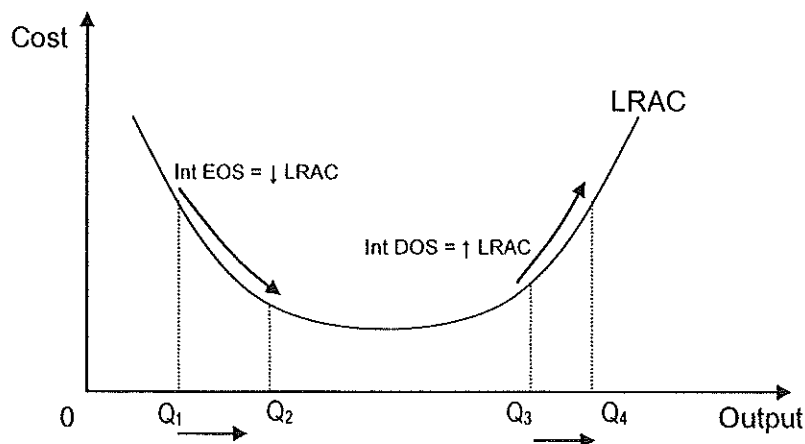
**e) Risk-bearing Diseconomies**

Even when a firm diversifies to reduce risk involved in production, it can prove unwise because when the firm diversifies too excessively, the poor performance of one branch of its operations may have negative spillover effects on all the other branches. So, inevitably the cost of production increases as output is increased.

**2.3.1.5.3 Economies and diseconomies of scale and effect on the LRAC**

- Internal economies of scale lowers cost per unit as the firm expands its output. It is represented graphically by a movement down the LRAC as shown in Diagram 13.
- Internal diseconomies of scale tend to raise the unit cost when the firm increases its output. Internal diseconomies is represented by a movement up the LRAC, as shown in Diagram 13.

**Diagram 13: Effect of Internal Economies & Diseconomies of Scale on LRAC**



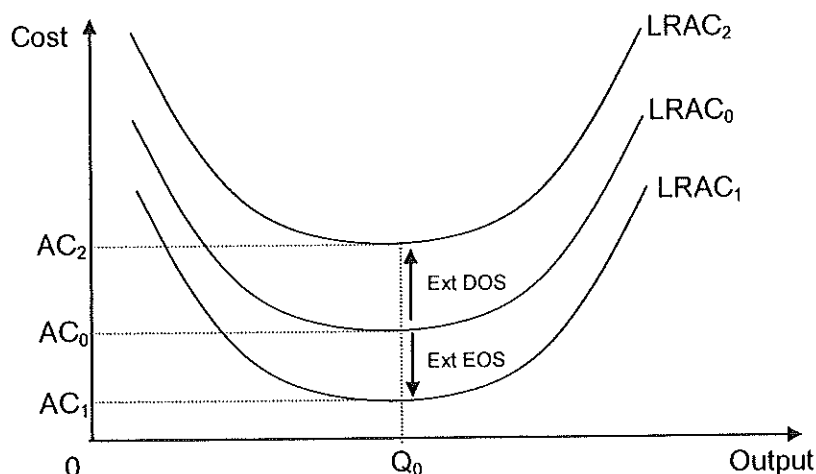
- The LRAC is normally U-shaped as it is assumed that as the firm expands its output, it experiences falling costs per unit arising from internal economies (represented by a movement down the LRAC curve) followed by rising costs resulting from diseconomies setting in (represented by a movement up the LRAC curve).



#### 2.3.1.5.3 External Economies of Scale:

- These refer to the reductions in unit cost enjoyed by a firm (regardless of its own individual size) as a result of the growth of the industry as a whole.
- External economies of scale cause a downward shift in the LRAC curve as average cost for each given output level is now lower. This is illustrated in diagram 14.

**Diagram 14:** Effect of External Economies & Diseconomies of Scale on LRAC



- When an industry grows in size and especially if the individual firms become concentrated in a particular place/area, the individual firms may enjoy benefits from the **geographical concentration**:
  - a) Development of infrastructure and support services**

A good transport and communication system may be set up to boost production and distribution. Each firm may have access to specialist raw material and component suppliers, firms which specialise in marketing the finished product and banks and other financial institutions which cater to the special needs of the industry. Local institutions may develop courses of training for the specialist skills required by this industry.
  - b) Development of research facilities**

The firms in the industry may jointly set up a research centre to bring about cost savings for all.
  - c) Development of firms supplying components**

When the industry grows in size, the production process may be broken up and each individual firm may specialise in the production of some component of the final product. Specialisation as such may lead to increases in productivity and cost savings would be achieved as each component may be obtained at a relatively low cost per unit as it is being mass produced for the industry.

#### 2.3.1.5.3 *External Diseconomies of Scale:*

- These cause an upward shift in the LRAC curve of the firm as a result of the industry becoming larger. It is illustrated by an upward shift of the LRAC as shown in Diagram 14. As the industry grows too large or firms concentrate in a particular area, external diseconomies tend to rise.
- a) **Rising input costs**  
Strong competition and rising demand for factors of production in that region may lead to rising prices of the factors of production, resulting from shortage of labour and land. Hence rising average costs per unit.
- b) **Overcrowding**  
Local road congestion may also cause transportation delays and transport costs to rise. Problems like traffic congestion, overcrowding, noise pollution and etc., will increase and may cause an increase in the overall cost of production.

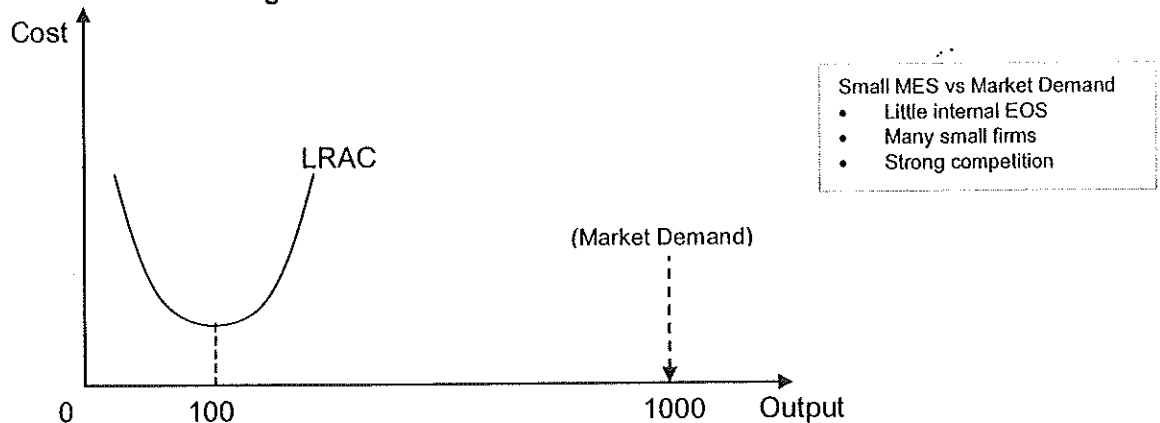
#### 2.3.1.6 *Size of Firms*

- There are many measures of the size of a firm. Some of the measures are:
  - number of employees
  - value of sales (sales turnover or total annual revenue)
  - value of fixed capital employed
  - market share out of total demand
  - others including tonnage for ships, acreage for farming.
- The measure used can differ from industry to industry. Small firms predominate in certain industries such as agriculture, retailing, building, personal and professional services. Even within an industry, there can be variations in the size of the firms.
- The firm can continue to remain in its present size or can expand over the years through gradual expansion or through acquisitions (takeovers) or mergers, exploiting most of the potential of economies of scale (see Appendix 1 for more details).

#### 2.3.1.7.4 Minimum Efficient Scale

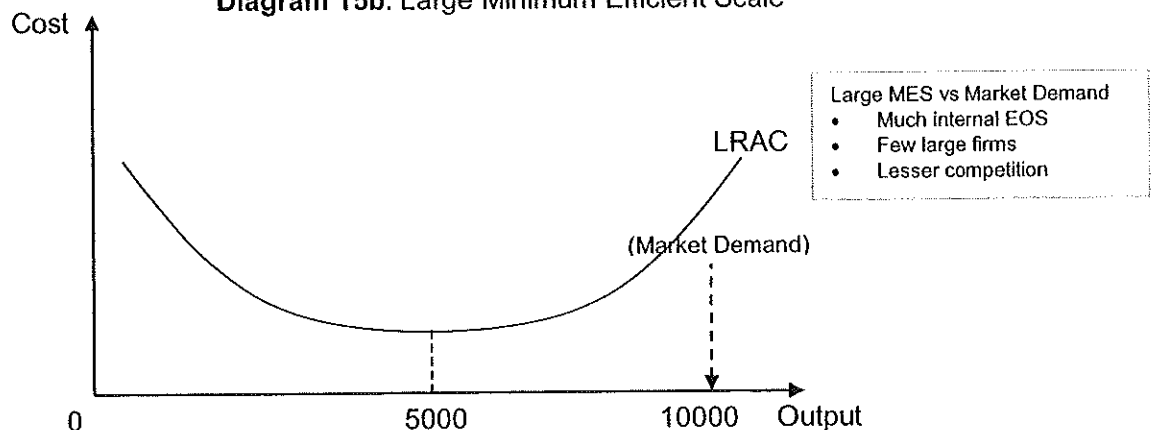
- **Minimum efficient scale (MES)** is the smallest output level at which the firm is able to minimise its long-run average cost. MES helps the firms in the industry to determine if it is worthwhile for them to grow or to remain small.

Diagram 15a: Small Minimum Efficient Scale

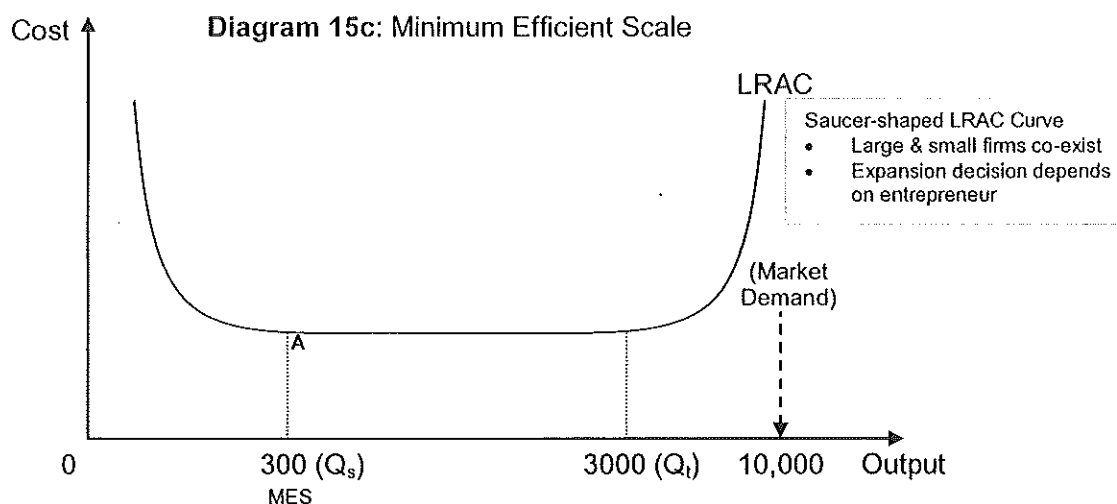


- Referring to Diagram 15a, MES is at 100 units of output.
- If MES is small relative to market demand, then firms in the industry are likely to remain small, as the scope for economies of scale is limited. As such, there are likely to be many small firms in the industry (e.g. retail industry). The degree of competition is likely to be high, as there is room for many efficient firms in the industry.

Diagram 15b: Large Minimum Efficient Scale



- MES is at 5000 units of output as shown in Diagram 15b.
- If MES is large relative to market demand, then the firms in this industry (e.g. oil refining) are likely to be large as there is substantial scope to enjoy economies of scale. Competition is likely to be lesser because there is room only for a relatively small number of efficient firms.



- MES is at point A (300 units of output) as shown in Diagram 15c.
- If firms in the industry face an LRAC like this, there is likely to be a combination of small and large firms in the industry. Larger firms will have no cost-savings over smaller firms, so long as the smaller firms have at least achieved MES. As such, whether the firm expands or not will depend on the objectives of the entrepreneur.

#### References :

Ellie Tragakes. Economics for the IB Diploma, Chapter 4. Cambridge University Press (2009), p91 – 112.

Blink J & Dorton I. Economic Course Companion. pg 73 – 94, IBO Oxford University Press, UK (2007)

Matt McGee. Economics – In terms of the Good, the bad and the Economist. Chapter 2.3, IBID Press, Australia (2009), p147 – 163.